STATEMENT OF PROJECT OBJECTIVES
DE-FC26-05NT42302

Design and Feasibility Assessment of a Retrospective Epidemiologic Study of Coal-Fired Power Plant Emissions in the Pittsburgh, Pennsylvania Region

A. OBJECTIVES

The overall objective of this project is to develop a strategy to better define the public health implications (both short term and, if possible, long term) in the Pittsburgh region of particulate matter emissions from coal-fired power plants, using retrospective epidemiological study methods applied to speciated $PM_{2.5}$ information, other emissions data, and other plausible explanatory factors.

In the process of achieving this overall objective, a succession of intermediate objectives will also be achieved, including:

- Assess the location, quantity, and quality of available speciated $PM_{2.5}$ air monitoring data (including archived data) in the Pittsburgh region from 1999 to 2004;
- Develop a plan including technical and economic evaluations for the construction of an air monitoring database for use in a retrospective epidemiology study;
- Assess the quality and quantity of available mortality and morbidity data (cardiopulmonary hospital admissions, emergency room visits, physician visits, etc.) in the Pittsburgh region from 1999 to 2004;
- Develop a plan including technical and economic evaluations for the construction of a health outcomes database for use in conducting this retrospective epidemiology study;
- Investigate and/or develop methods to apportion $PM_{2.5}$ constituents to sources;
- Employ the use of a geographic information system (GIS) to spatially model exposure due to traffic density and proximity to roadways;
- Explore, evaluate, compare, and validate the use of various statistical modeling techniques to help control for potential confounding of any coal-fired power plant $PM_{2.5}$-related constituents and morbidity/mortality association by factors such as other sources of $PM_{2.5}$ constituents, co-pollutants, weather, traffic density, and proximity to roadways;
- Determine whether a successful retrospective (1999-2004) epidemiology study modeling specific health outcomes as a function of source-related $PM_{2.5}$ constituents and other confounding factors in the Pittsburgh region is feasible.

B. SCOPE OF WORK

This project involves three distinct but interdependent main tasks:

Task 1 - Inventory and assessment of air monitoring data - A plan will be developed by completing a series of subtasks during the first 6 to 7 months of the performance period. The subtasks include inventorying existing air monitoring data, assessing the quality of these data, determining whether additional data can be acquired, and assessing the feasibility of using the data in a retrospective epidemiological study.

Task 2 - Inventory and assessment of health impact data - A series of subtasks will be completed during the first 6 months of the performance period in order to evaluate the feasibility of a retrospective epidemiological study and to develop a broad plan for the most efficient study design. The subtasks include inventorying existing health outcomes databases, assessing the quality of the available data, investigating whether additional data can be acquired through existing retrospective records and cohorts, and deriving an overall assessment of the utility of the data for use in a retrospective epidemiological study.

Task 3 - Comprehensive evaluation and selection of possible methods for the statistical analysis of the data - Concurrently with Tasks 1 and 2, a plan will be developed to use GIS and exploratory data analysis techniques to provide a comprehensive overview of the region, to investigate various possible methods for dealing with multiple monitoring sites and for spatial/temporal aggregation of emission data and other explanatory factors to avoid exposure misclassification, to provide source apportionment of $PM_{2.5}$ emissions, and to provide effects modeling and estimation based on available data from the Pittsburgh region and other regions.

The project team also intends to investigate the use of local retrospective cohorts assembled previously for various national and local research studies such as the Cardiovascular Health Study, Health ABC, Women's Health Initiative, etc. or constructed from data gathered from health maintenance organizations (HMOs) and area physicians for the evaluation of the longer-term (months and longer vs. weeks and less) effects of air pollution on human health endpoints. This requires investigation of methods useful for constructing a plausible model for observed effects as a function of emissions having controlled for as many potential confounding factors as possible, including factors concerning the spatial locations and temporal availability of monitoring data for short term effects. Given the observational nature of retrospective investigations, it is impossible to control for all potential confounding factors but the credibility of the assessment will be in proportion to handling appropriately as many plausible confounders as possible using the existing database. The complexity of the possible methods of analysis is at least partly due to restrictions engendered by a retrospective and observational study.

C. TASKS TO BE PERFORMED

This section contains detailed descriptions of major tasks and subtasks to be performed during the project.

Task 1 - Inventory and assessment of air monitoring data - In order to determine the feasibility of conducting a retrospective $PM_{2.5}$ epidemiology study with source apportionment in the Pittsburgh region, a thorough assessment of the air monitoring data available for the region is required. There are four (4) subtasks to be performed under Task 1:

> Task 1.1 - Inventory available $PM_{2.5}$, gaseous pollutant, and meteorological data. A comprehensive inventory of air monitoring data available for the Pittsburgh region for the period 1999 through 2004 will be assembled. Potential sources of speciated $PM_{2.5}$, gaseous pollutant, and/or meteorological data include the Upper Ohio River Valley Project, Pittsburgh Air Quality Study, DOE-NETL in-house monitoring program, Allegheny County Health Department, Steubenville Comprehensive Air Monitoring Program, and Pennsylvania DEP. In addition to thoroughly surveying data that are available in the $PM_{2.5}$ Ambient Fine Particulate Research Air Quality Database,

representatives from each of these air monitoring efforts will be contacted to determine, by day and location, what speciated $PM_{2.5}$ data, gaseous pollutant data, and meteorological data are available for Pittsburgh and surrounding areas between 1999 and the present time. Compositional data of interest include concentrations of SO_4^{2-}, NO_3^-, elemental carbon, organic carbon, and various elements (e.g., Se, Mn, Zn, Fe, V, Ni, Al, K, Si, Pb, Cu); gaseous pollutant data of interest include concentrations of CO, NO_x, SO_2, and O_3; and meteorological data of interest include temperature, relative humidity, barometric pressure, wind speed, and wind direction. Sampling, analytical, and QA/QC procedures used to obtain the data will be documented, and periods of missing or invalid data will be noted. Inventory results will be summarized in a grid showing, with a daily time resolution, what data are available for each monitoring site in the Pittsburgh region.

Task 1.2 - Assess the quality and comparability of available air monitoring data. Air monitoring data that are available for a retrospective epidemiology study in the Pittsburgh region were collected by different groups who used different measurement techniques and followed different QA/QC protocols. Hence, an assessment of the quality and comparability of the available data is necessary. For example, in order to determine the elemental composition of $PM_{2.5}$, UORVP and ACHD used x-ray fluorescence (XRF), PAQS used conventional inductively coupled plasma - mass spectrometry (ICP-MS), SCAMP used dynamic reaction cell ICP-MS (DRC ICP-MS), and DOE-NETL used proton induced x-ray emission (PIXE). The accuracy and precision of these measurement techniques can vary appreciably, especially for certain key elements. Conrad et al. (2000) reported that results obtained using DRC ICP-MS and XRF generally agree within 20% for Se and Pb, within 30% for Fe and Mn, and within 30%-70% for Ca, Zn, K, and Al. Si concentrations determined using the two methods generally differ by a factor of 5 to 10. In addition, XRF lacked the sensitivity to accurately determine most other trace elements. Hence, prior to specifying data for possible use in source apportionment or in an epidemiological model, the uncertainties and comparability of these data will be assessed. If necessary, the possibility of applying correction factors to improve the comparability of data collected using different measurement techniques will be investigated. Additionally, QA/QC procedures followed by each monitoring program will be reviewed to determine what, if any, additional data reduction steps would be required to ensure that consistent data validation rules are applied to all data specified for use in a possible future epidemiology study.

Task 1.3 - Inventory archived filters that might be analyzed to augment the speciated $PM_{2.5}$ data record. It is known or believed that some $PM_{2.5}$ samples collected as part of the UORVP, DOE-NETL, ACHD, and SCAMP monitoring efforts are currently archived, and may be available for future analysis (with permission from the project group and project sponsor). Determination of the chemical composition of these samples could substantially augment the speciated $PM_{2.5}$ data record for the Pittsburgh region between 1999 and the present time. Hence, site visits will be performed to assess the condition and availability of archived filters from each of the aforementioned monitoring programs. The feasibility of performing laboratory analyses on these archived filters to obtain sufficiently detailed chemical speciation data for the purposes of an epidemiology study with source apportionment will also be assessed. Methodological issues identified in Task 1.2 that are related to data comparability will be considered, as will practical limitations of available analytical techniques. For instance, if only a single filter is available for a given day at a given location, a non-destructive method such as XRF may be preferable to a destructive method such as DRC ICP-MS for the determination of elemental composition; however, XRF may not have sufficient sensitivity to determine concentrations of key species such as Zn, Si, or V. If only Teflon filters are available for a given site, thermal optical transmittance (TOT) and thermal optical reflectance (TOR) methods that are typically used to determine concentrations of elemental and organic carbon cannot be used. An optical absorption technique may be employed to determine the concentration of light absorbing (elemental) carbon; however, this method does not determine the concentration of organic carbon, and there will likely be substantial uncertainty associated with the elemental carbon determination. Based upon considerations such as these, analytical procedures for determining $PM_{2.5}$ composition for archived filters will be recommended, and the cost and time that would be required to perform these analyses will be estimated. Findings concerning additional $PM_{2.5}$ compositional data that can be obtained from the analysis of archived filters will be added to the grid described under Task 1.1.

Task 1.4 - Develop a plan for the construction of an air monitoring database for use in a retrospective epidemiology study of $PM_{2.5}$ and its components. Based upon the findings in Tasks 1.1 through 1.3 as well as epidemiological considerations (e.g., statistical power, definition of geographical region, Tasks 2.1 through 2.4) and source apportionment requirements (e.g., tracer species to be used for source apportionment, Tasks 3.1 through 3.3), a plan will be developed for the construction of an air monitoring database that is suitable for use in the proposed epidemiology study. The plan will include a technical evaluation detailing: (1) which currently available data should be included in the database, as well as what, if any, additional data reduction work is required to make these data suitable for analysis, (2) which archived $PM_{2.5}$ filters should be analyzed to determine $PM_{2.5}$ chemical composition, as well as the analytical techniques that should be used to perform these analyses, and (3) what, if any, additional ambient air monitoring is needed to permit a retrospective $PM_{2.5}$ epidemiology study, including a description of monitoring location(s), duration, time resolution, and methodologies. The plan will also include an economic evaluation. Development of the plan essentially amounts to solving a constrained optimization problem – the plan must minimize expenditures of time and money while achieving a quantity and quality of air monitoring data sufficient to enable the performance of a retrospective epidemiology study of $PM_{2.5}$ and its components in the Pittsburgh region.

Task 2 - Inventory and assessment of health impact data - In order to determine the feasibility of conducting a retrospective $PM_{2.5}$ epidemiology study with source apportionment in the Pittsburgh region, a thorough assessment of health outcomes data available for the region is required. There are three (3) subtasks to be performed under Task 2:

Task 2.1 - Assess the quality and quantity of information within the existing health outcomes (mortality and morbidity) databases available for the required time period (1999-2004). A critical component of any retrospective epidemiological study of $PM_{2.5}$ and health outcomes is the identification of readily available and accessible mortality and morbidity databases for the local region. Given the trends in improved treatments for disease, mortality alone is not a sensitive enough indicator to capture any potential effect of daily changes in air pollution on health. Ideally, daily or even hourly medical information would be available to capture all health-related outcomes in the population potentially related to variations in $PM_{2.5}$ concentrations and/or its components. For the time period of interest, this information would include but not be limited to deaths, hospital admissions, emergency room visits, physicians' office visits, prescriptions, and medication use, all preferably in electronic format. Although the level of detail required for an ideal comprehensive assessment will most probably not be available in retrospective datasets, health outcomes in Pennsylvania and the Pittsburgh region are captured by several data collection entities, some of which are unique to this region.

From previous work and a preliminary evaluation, several databases are known to be available from 1999 to 2004 that might be used to reconstruct the health outcomes profiles of residents in the Pittsburgh region. For mortality, potential resources include the Allegheny County Health Department (ACHD), the Pennsylvania Department of Health Statistics, and the Ohio Department of Health. For morbidity, daily hospital admissions might be the most well-defined and accessible estimate of the health effects potentially related to $PM_{2.5}$ and its components. These data are systematically collected for the region and are available retrospectively via the Pennsylvania Health Care Cost Containment Council (PHC4) and the Ohio Hospital Association. Information on daily emergency room visits are available from: the Medical Archival System (MARS), a proprietary software system of the University of Pittsburgh Medical Center (UPMC); the Real-time Outbreak Disease Surveillance (RODS) Laboratory, a University of Pittsburgh real-time computer-based public health surveillance system; and individual local and regional hospital databases.

To investigate long-term morbidity and mortality effects, the study team will investigate the feasibility of entering into an agreement with local Health Maintenance Organizations (HMOs) for restricted access to de-identified health care data of subscribers. Data from local population-based cohorts assembled in past years for national and regional research studies such as the Cardiovascular Health Study (CHS), Women's Health Initiative (WHI), Study of Women Across the Nation (SWAN), MR. FIT and others represent another potential resource for evaluating the health effects of long-term exposure to fine particulates. Medicare billing information, the Scott-Levin pharmaceutical use database, UPMC Health Plan Pharmaceutical Database, and others will also be tapped as possible resources for a retrospective analysis, particularly in more susceptible age groups (i.e. Medicare-65 years and older; pharmaceutical databases-children).

<u>Task 2.2 - Assess the availability of data related to population density, traffic density, proximity to roadways, mileage of paved roads, etc</u>. The Federal Bureau of Transportation, the Pennsylvania Department of Transportation (PENNDOT), the RAND Corporation and other groups collect municipality-specific data on traffic volume, mileage of paved roadways, etc. Certain information can be collected from the traffic engineering files of these agencies. PENNDOT has implemented an on-line system for the public to explore GIS- based, municipality–level traffic data such as vehicles per day, etc. GIS software packages such as ArcInfo (ESRI Inc) provide high resolution road network data. In addition, several companies such as Geographic Data Technology (NH) provide road network databases for research and other purposes. Various municipality-level population based indices, including population densities, are available through U.S

Census data and GIS (ArcView). All available datasets will be explored for suitability for retrospective analysis from 1999-2004.

Task 2.3 - Develop a plan for the construction of specific and comprehensive health outcomes databases for use in a retrospective epidemiology study of $PM_{2.5}$ and its components. Based upon the findings in Tasks 2.1 and 2.2 as well as epidemiological considerations (e.g., statistical power, definition of geographical region), a plan will be developed for the construction of several independent as well as a comprehensive health outcomes databases that are suitable for use in the proposed epidemiology study. The plan will include a technical evaluation detailing: (1) which currently available data should be included in the database, as well as what, if any, additional data reduction work is required to make these data suitable for analysis, (2) what, if any, additional health outcomes information is needed to permit a retrospective $PM_{2.5}$ epidemiology study. The plan will also include an economic evaluation. Development of the plan essentially amounts to solving a constrained optimization problem – the plan must minimize expenditures of time and money while achieving a quantity and quality of air monitoring data sufficient to enable the performance of a retrospective epidemiology study of $PM_{2.5}$ and its components in the Pittsburgh region.

Task 3 - Comprehensive evaluation and selection of possible methods for the statistical analysis of the data - There are a number of important issues concerning statistical methodology that must be correctly overcome when attempting to relate health effects to coal-fired power plant emissions and other pollutants. These issues include, among others, effective measures to remove confounding due to other causal agents such as emissions from other pollution sources and effects due to meteorology, modeling and/or removing serial dependence (autocorrelation) in each time series, correctly apportioning emissions to sources, and determining exposure from multiple monitoring sites to avoid error due to exposure misclassification. The choices made to address one area may influence choices made in other areas. There are four (4) major subtasks to be performed for Task 3:

Task 3.1 - Perform an exploratory data analysis and create a GIS Platform for the Pittsburgh Region. The use of geographical information system (GIS) methods along with exploratory data analysis methods will be investigated as ways to facilitate understanding of emission sources, and of the regional trends and variability of pollutants in the Pittsburgh region using existing databases available from the Allegheny County Health Department and others.

Detailed land use, land cover and demographic mapping data will be integrated and analyzed on a GIS platform for the Pittsburgh region. This platform will incorporate an extensive disparate data set to assess the feasibility of developing estimates for the spatial distribution of $PM_{2.5}$, its components and co-pollutants. The analysis will also aim to provide a measure of uncertainty for these estimates. The spatial distribution of the pollutants will be linked to different geographic attributes such as demographics, traffic density, industrial zones and mortality/morbidity data. In addition the feasibility of using this platform to identify susceptible populations potentially attributed to areas of influence from specific sources in the Pittsburgh region such as high density traffic zones and point sources will also be investigated. The GIS platform will incorporate data related to 1) population density/demographics; 2) traffic density; 3) point source emissions; 4) morbidity and mortality; 5) ambient monitoring of pollutant concentrations derived from DOE sponsored programs; and 6) ambient monitoring of pollutant concentrations from state, federal and local air agencies.

These methods would also be especially useful for determining the optimum geometric boundaries for the region to be studied and for comparisons between regions with differing emission levels for the examination of long term effects.
Investigate, test, compare and contrast exploratory data methods along with the use of GIS that would provide insight into the spatial and temporal trends and variability of health outcomes, $PM_{2.5}$ components and other explanatory variables in preparation for source apportionment and the use of effects modeling techniques.

Task 3.2 - Compare individual site information versus spatial/temporal averaging of emissions and explanatory variables. The various air monitoring programs conducted in the Pittsburgh region between 1999 and 2004 collected data at a number of geographically distinct locations; however, these monitoring programs were conducted at different periods in time. Hence, although some time periods feature data collected at multiple monitoring sites, other periods likely feature data collected at only a single site. It is thus unlikely that a multiple-year, spatially resolved data stream can be constructed for the Pittsburgh region using only data and $PM_{2.5}$ samples that have already been collected. Rather, a single "number" may have to be used to represent the concentration of each $PM_{2.5}$ constituent and gaseous pollutant and the value of each weather variable for the entire region for each day of the study.

In order to assess the short-term health impact of various sources of fine particulates, it would ideally be required to know the exposure to fine particulates from coal-fired plants for each individual in the region of interest along with the health outcome for that individual at regular time points (for example, daily) and all potential confounding variables (weather and other pollutants and personal characteristics such as smoking behavior) over a long period of time for many individuals. Health effects could be indicator variables (hospitalized or not hospitalized on a given day, for example) or some continuous measure of health. Time series models for each person would estimate the effect on health of changing levels of the emissions while filtering out the effects of all confounding factors. The individual time series could be appropriately combined or aggregated. If the health effect were an indicator variable, the aggregate would then be a count (for example, the number of hospitalizations on a given day).

Actual individual exposures and corresponding health outcomes will not be known so it will be necessary to try to approximate this ideal information in an optimal manner. The available information will consist of speciated $PM_{2.5}$ concentration measurements and other explanatory variables (for example, weather related variables and other pollutants) made by a small network of monitoring sites along with aggregate counts of mortality and morbidity, and possibly community events related to health, e.g., influenza epidemics, all on a daily basis (or converted to a daily basis). There will be no comprehensive information on the health of individuals, the location of the monitoring sites will change over time and there will be periods of time with no daily information on one or more (or maybe all) variables. It is clear then that these data could not be used directly to approximate the information on separate individuals if only because the health outcomes for individuals will not be known. With individual time series models for each person, we would be able to connect individual exposures to health incomes. This would provide information on health effects both longitudinally (comparisons between different time points for a given individual) and cross-sectionally (person-to-person comparisons). Because the connection between actual exposure and health outcome is missing, the evidence for short-term health effects must necessarily come solely from the longitudinal information (changes in emissions over time) for the population as a whole.

For example, if the health outcomes are aggregate daily counts (or averages of other continuous health outcomes), then the explanatory variables (exposure information broadly defined and the measurements on confounding variables) may need to be aggregated also. To do this in an optimal manner may require trying to preserve as much of the connection between the emissions data and other explanatory variables and the individuals as possible before aggregation takes place. What assumptions are most reasonable to make and how best to make the estimates will be investigated. If, for example, health outcome data are available with some location information (for example, say by zip code) then it might make sense to investigate methods of spatial statistics (a.k.a. geostatistics) to provide regionalized estimates of exposure. These regionalized exposure estimates might provide a more realistic index for individual exposure or at least for exposure in the local subregions. These subregions could then be aggregated to gain statistical power.

It should be kept in mind that after the region of interest is finalized (and the choice of the size and shape of the region probably depends to some extent on what monitoring sites are available), even measurements taken outside the delimited region may be used to inform the spatial/temporal averages (and subregional averages, if appropriate). Whether outside sites are useful will depend in part on the outcome of any spatial statistical analysis.

<u>Task 3.3 - Apportion emissions to sources and evaluate the results</u>. In order to determine health effects due to coal-fired power plants, it will be necessary to determine some overall measure of the emissions due solely to coal-fired power plants and also to each other major source, including motor vehicles, local sources such as steel mills, coke ovens, and chemical plants, and possibly others. The ideal approach would include having a complete and exhaustive inventory of the emissions from each coal-fired plant, and for every other source, within the delimited region and those outside the delimited region that affect the region. Information on emission sources will be inventoried.

Source apportionment ideally creates a chemical mass balance (CMB) between sources and receptors. The information is usually not available when dealing with a large region over a long period of time. Multivariate statistical techniques, in particular, factor analysis, have been used to determine source apportionment in many studies.

There are four major limitations of exploratory factor analysis (EFA) methods:

- Cannot incorporate constraints on the source profile matrix needed to obtain a physically valid solution.
- Factor indeterminancy prevents a unique and interpretable solution.

- Does not incorporate partial source profile information.
- Temporal dependence in data results in invalid statistical inference.

Exploratory factor analysis is used when there is little physical and chemical knowledge available to incorporate directly into the model. EFA can be used to “discover” potential sources (of unknown number) but does not address the four limitations listed above. Positive matrix factorization (PMF) and UNMIX are two examples of using constraints to obtain physically valid solutions without knowing the number of sources but they only address the first limitation. Confirmatory factor analysis (CFA) uses some pollution source information, hypothesizes and tests for the number of factors, and allows various constraints based on knowledge of source emissions to be imposed to limit the number of possible solutions. CFA addresses the first three limitations. Bayesian factor analysis (BFA) goes one step farther and places priors on the elements of the source profile matrix. Going farther still, regression and measurement error models are used when both the number of sources and the source profile matrix are known. None of these approaches address the fourth limitation of serial dependence of the observation vectors. To address the lack of serial independence, the use of nested block bootstrapping to provide parameter estimates and valid statistical inferences may be helpful.

These and related methods will be investigated and tried on test data sets. The results of these modeling efforts will be used to transform the observed PM2.5 concentrations into source scores (e.g., transportation, coal-fired plants, industry, and so forth). The source scores will be used in place of the $PM_{2.5}$ concentrations in the empirical models.

Task 3.4 - Develop and evaluate statistical methods for modeling health effects as a function of coal-fired power plant source emissions while controlling for potential confounding explanatory factors. Using individual monitoring site information and/or regional averages for each explanatory variable, the goal will be to develop and evaluate statistical modeling methods for relating the health impact counts to the coal-fired emission sources while controlling for all the other explanatory variables. Whatever modeling technique is used, it will be necessary to ensure that, to the extent possible, all model assumptions are reasonable and, if at all possible, to validate the method by realistically comparing model predictions with reality. To do the evaluations and develop methodology, it will be necessary to use easily available data from the Pittsburgh region and possibly other regions.

While confounding by other causal factors is extremely important, the handling of autocorrelation in the measurements must also be adequately addressed through appropriate modeling techniques and assessed through appropriate model diagnostics.

The pros and cons of various modeling strategies will be investigated. There are at least three principal approaches to modeling environmental time series data to look for short term effects: 1) generalized additive models (GAM), 2) generalized linear models (GLM, and extensions such as generalized ARIMA or GARIMA and generalized linear autoregressive moving average or GLARMA), and 3) state space models which easily handle irregularly spaced time points and missing values. These models differ in terms of how they handle seasonal effects, trends and confounded explanatory factors.

GAM and GLM have been extensively used in modeling the health effects of particulates. It may be useful to compare the results of these modeling efforts to state space models and related methodologies so that the optimum approach can be used. These methods differ in their approaches to handling seasonality and autocorrelation.

The use of generalized additive models (GAMs) in some studies has resulted in effects estimates being biased upward along with unrealistically low standard errors even when the computations properly converge. Modifications to GAM estimation will be investigated as will generalized linear models (GLM - a parametric modeling approach). State space time series methodology is attractive because it encompasses a wide range of models and has been applied to many problems, the computations are well-understood, and the results are robust.

Project Schedule

Figure 1 depicts the overall schedule for project activities. Specific start dates and durations for subtasks are not provided, because these specific activities are anticipated to occur intermittently throughout the time period covered by the major task.

Figure 1 – Schedule of Project Activities, DE-FC26-05NT42302

		Months after Start of Project											
Task	**Description**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1	**Inventory and assessment of air monitoring data**												
1.1	Inventory of PM2.5, gaseous pollutant, and meteorological data.												
1.2	Assess quality and comparability of data												
1.3	Inventory archived filters												
1.4	Develop plan for air monitoring database												
2	**Inventory and assessment of health impact data**												
2.1	Assess existing health outcomes databases												
2.2	Assess data related to population density and traffic-related parameters												
2.3	Develop plan for health outcomes database												
3	**Evaluation and selection of statistical analysis methods**												
3.1	Create GIS Platform												
3.2	Compare individual site information versus spatial/temporal averaging												
3.3	Apportion emissions to sources; evaluate results												
3.4	Develop and evaluate statistical methods												

Roles of Participants

University of Pittsburgh, Graduate School of Public Health, Department of Epidemiology will assess the availability and quality of existing health outcomes and population-based data, and the statistical analysis. Jeanne Zborowski will oversee and conduct, with the assistance of a project director, the activities outlined in Task 2. Evelyn Talbott will support activities with primary emphasis on developing the plan for the comprehensive health outcomes databases (Task 2.3) and designing the retrospective epidemiological study (Task 2.3). Additional support may be required from the Department of Epidemiology project group to support database activities in Task 1. Rick Bilonick and

Nancy Sussman will oversee and conduct the statistical analysis activities for Task 3, and with Evelyn Talbott oversee the development of an overall strategy. Larry Keller will work with CONSOL R&D on Task 1.

CONSOL Energy R&D will assess the availability and quality of existing air monitoring data. Dan Connell will oversee and conduct the activities outlined in Task 1. Steve Winter will support activities with primary emphasis on Task 1.2, assessing the quality and comparability of available air monitoring data, and Task 1.3, conducting an inventory of archived filters that might be analyzed to augment the speciated $PM_{2.5}$ data record for the Pittsburgh region. Additional support may be required from CONSOL Energy R&D's systems group to support programming and database activities in Task 1. Steve Winter and Dan Connell will also work with Rick Bilonick and Nancy Sussman on exploratory data analysis, handling data from multiple monitors, and source apportionment, Tasks 3.2 and 3.3.

Allegheny County Health Department will assist CONSOL R&D. Sam Schlosberg will work with Steve Winter, Dan Connell, and Larry Keller (GSPH) on Task 1.

Ohio University will provide expertise on GIS modeling. Kevin Crist will oversee and support the work of JB Hoy and Darren Cohen, primarily on Tasks 2.2 and 3.1.

D. DELIVERABLES

A project kickoff meeting will be held within 60 days after project award. The project team will prepare a presentation summarizing the objectives and work to be performed during the project, and distribute this presentation to all project participants to facilitate discussion during the kickoff meeting.

On a quarterly basis, Federal Assistance Program/Project Status Reports and Financial Status Reports will be prepared and submitted to DOE/NETL.

On a quarterly basis, Technical Progress Reports will be generated to summarize the results of the project. Each Technical Progress Report will include a summary of all data obtained, problems encountered, and plans for the immediate future. It is anticipated that the Technical Progress Reports will contain the following information, as it is developed:

- Inventory of currently available air monitoring data - The grid will summarize the available data for each monitoring site in the Pittsburgh region. The grid will be qualified for deficiencies, QA/QC, measurement considerations, and overall utility.
- Inventory of archived filters – The inventory will identify archived filters that could be analyzed to provide additional speciation data. These data may be required to generate a time resolved data stream of sufficient duration that will contribute to the statistical power necessary for the epidemiology study.
- Plan for construction of an air monitoring database – The plan will consider all the technical issues and the economics to develop a database for use in retrospective epidemiology study of $PM_{2.5}$ and its components in the Pittsburgh region.
- Inventory of existing morbidity and mortality databases- The study team will provide a comprehensive inventory of all existing health outcomes databases for the Pittsburgh

region that cover the period from 1999-2004, including a summary of variables contained in each database and strengths and weaknesses of the available dataset.

- Inventory of available data related to population density and traffic indicators - The study team will provide a comprehensive inventory of all existing databases related to traffic indicators for the Pittsburgh region that cover the period from 1999-2004, including a summary of variables contained in each database and strengths and weaknesses of the available dataset.
- Comprehensive plan, including technical and economic evaluations, for the construction of specific and comprehensive health outcomes databases for use in a retrospective epidemiology study of $PM_{2.5}$ and its components in the Pittsburgh region.
- A comparison of statistical methods used to describe and explore data trends and variability, aggregate data to provide consistent regional averages for each explanatory variable, to apportion daily regional emission averages to sources, and to construct models correctly accounting for confounding factors based on similar relevant data sets in terms of robustness and sensitivity of results to particular methodologies. Potential methods of model validation will also be tested.
- An evaluation of the applicability of the statistical methods for the prospective database for the Pittsburgh region to be assembled in a future study; and, if appropriate, a plan to apply these methods.
- Comprehensive assessment of the feasibility and preliminary design of a retrospective epidemiological study of the relationship between $PM_{2.5}$ and its components and health outcomes in the Pittsburgh region - The feasibility of the study will be determined by the quantity, quality and availability of air monitoring and health outcomes data for the Pittsburgh region from 1999-2004. The study design will incorporate the evaluation of both short- and long-term effects of exposure.

A final report will be issued at the end of the program summarizing the results of the feasibility assessment. Environmental reports will be prepared, including a Hazardous Substance Plan once the award is made and a Hazardous Waste Report at the end of the program. A Report of Termination or Completion Inventory will also be submitted at the end of the program.

Table 1 is a schedule of key deliverables/milestones associated with the project, based on a nominal start date of March 18,, 2005.

Table 1 - Milestones and Deliverables, DE-FC26-05NT42302

Milestone/Deliverable	Date
Project Kickoff Meeting	04/01/05
Quarterly Progress Reports	07/18/05; 10/18/05; 01/18/06
Draft Final Report	04/18/06
Final Report	06/18/06

E. BRIEFINGS/TECHNICAL PRESENTATIONS

A project kickoff meeting will be held within 60 days after project award. The project team will prepare a presentation summarizing the objectives and work to be performed during the project, and distribute this presentation to all project participants to facilitate discussion during the kickoff meeting.

It is anticipated that descriptions and/or results of the project will be presented at epidemiology conferences, workshops, and other public meetings. If participation in such events will be supported by the Cooperative Agreement, the project team will provide information (dates, location, etc.) to the DOE COR prior to the event.

In addition, the project team will present all findings to DOE management and technical personnel as requested by the DOE COR.